<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>A Weibull-count approach for handling under- and overdispersed longitudinal/clustered data structures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Luyts, Martial; Molenberghs, Geert; Verbeke, Geert; Matthejs, Koen; Ribeiro Jr, Eduardo E.; Demétrio, Clarice G. B.; Hinde, John</td>
</tr>
<tr>
<td><strong>Publication Date</strong></td>
<td>2018-07-30</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>SAGE Publications</td>
</tr>
<tr>
<td><strong>Link to publisher's version</strong></td>
<td><a href="https://doi.org/10.1177/1471082X18789992">https://doi.org/10.1177/1471082X18789992</a></td>
</tr>
<tr>
<td><strong>Item record</strong></td>
<td><a href="http://hdl.handle.net/10379/15792">http://hdl.handle.net/10379/15792</a></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
<td><a href="http://dx.doi.org/10.1177/1471082X18789992">http://dx.doi.org/10.1177/1471082X18789992</a></td>
</tr>
</tbody>
</table>
A Weibull-count approach for handling under- and overdispersed longitudinal/clustered data structures

Martial Luyts\textsuperscript{1}, Geert Molenberghs\textsuperscript{1}, Geert Verbeke\textsuperscript{1}, Koen Matthijs\textsuperscript{2}, Eduardo E. Ribeiro Jr\textsuperscript{3}, Clarice G. B. Demétrio\textsuperscript{3}, and John Hinde\textsuperscript{4}

\textsuperscript{1} Interuniversity Institute for Biostatistics and Statistical Bioinformatics, KU Leuven and/or Universiteit Hasselt, Leuven and/or Diepenbeek, Belgium
\textsuperscript{2} Family and Population Studies, KU Leuven, Leuven, Belgium
\textsuperscript{3} ESALQ, Piracicaba, University of São Paulo, São Paulo, Brazil
\textsuperscript{4} School of Mathematics, Statistics and Applied Mathematics, NUI Galway, Galway, Ireland

\underline{Address for correspondence:} Martial Luyts, L-Biostat, KU Leuven, Kapucijnenvoer 35, B-3000 Leuven, Belgium.
\underline{E-mail:} martial.luyts@kuleuven.be.
\underline{Phone:} (+32) 16 37 24 03.
\underline{Fax:} (+32) 16 33 70 15.

\underline{Abstract:} A Weibull-model-based approach is examined to handle under- and
overdispersed discrete data in a hierarchical framework. This methodology was first introduced by Nakagawa and Osaki (1975), and later examined for under- and overdispersion by Klakattawi et al. (2018) in the univariate case. Extensions to hierarchical approaches with under- and overdispersion were left unnoted, even though they can be obtained in a simple manner. This is of particular interest when analyzing clustered/longitudinal data structures with dispersion, where the underlying correlation structure is often more complex compared to cross-sectional studies. In this paper, a random-effects extension of the Weibull-count model is proposed and applied to two motivating case studies, originating from the clinical and sociological research fields. A goodness-of-fit evaluation of the model is provided through a comparison of some well-known count models, i.e., the negative binomial, Conway-Maxwell-Poisson, and double Poisson models. Empirical results show that the proposed extension flexibly fits the data, more specifically, for heavy-tailed, zero-inflated, overdispersed and correlated count data. Discrete left-skewed time-to-event data structures are also flexibly modelled using the approach, with the ability to derive direct interpretations on the median scale, provided the complementary log-log link is used. At last, a simulation study is conducted to examine other characteristics such as computational ease and orthogonality properties of the model, with the conclusion that the approach behaves best for highly overdispersed cases.

**Key words:** Longitudinal profiles; Clustering; Dispersion; Random-effects; Weibull-count approach
1 Introduction

The analysis of count data has received considerable attention in the literature, with practical applications in public health, and social and behavioral sciences. Since the introduction of generalized linear models (GLM’s) by Nelder and Wedderburn (1972), a GLM based on the Poisson distribution, a well-known member of the exponential family, is a commonly applied statistical model towards count data analysis. In spite of many advantages, e.g., the ability of fitting skewed non-negative data, the model possesses a too restricted mean-variance relationship (equidispersion), a characteristic that is often violated in the data. In particular, two situations can occur, (1) the variability in the data is larger than the theoretical variance function implied by the model (overdispersion), and (2) the variability in the data is smaller than the theoretical variance (underdispersion). For these and other reasons, e.g., zero-inflation (Iddi and Molenberghs, 2013) and heavy-tailed profiles (Zhu and Joe, 2009), many alternative and extended models have been proposed in the literature.

These models can often be classified as exponential dispersion models (EDM’s), introduced by Jørgensen (1987), which include the GLM families as a special case. More specifically, EDM’s increase the range of univariate/multivariate variance functions for which generalized linear type models exist. Kokonendji et al. (2004), for example, investigated two classes of EDM’s for count data that is overdispersed compared to the Poisson distribution, i.e., the Poisson-Tweedie and Hinde-Demétrio classes. Efron (1986), on the other hand, proposed a different class of regression families that include a model for dispersion, by introducing a second parameter in the exponential family that controls the variance independently of the mean while still carrying out the usual regression analysis in a GLM context. These are the so-called double-
exponential families because they enjoy exponential family properties simultaneously for the mean and dispersion parameters. A popular member is the double Poisson (DP) model (Appendix A4). A general overview of some popular models are given in Appendix A for subsequent comparison (Sections 4 & 5).

While most of these models find their origin back in the Poisson GLM framework, alternative approaches for modeling count data based on time-to-event distributions have recently been developed. These approaches are mainly built upon the direct relationship between the Poisson and exponential distributions (Cooper, 2005). Zeviani et al. (2014), for example, focused on a discrete version of the Gamma distribution to model count data by following the two-step approach of Winkelmann (1995): (1) Define the Poisson process as a sequence of i.i.d. exponentially distributed waiting times (Cox, 1962); and (2) replace the exponential distribution with a less restrictive (extended) non-negative distribution such as the Gamma distribution. For the Weibull distribution, Morais and Barreto-Souza (2011) constructed count versions, i.e., the generalized Weibull power series (GWPS) class of distributions. Another, yet simple approach based on the Weibull distribution, is the discrete approach of Nakagawa and Osaki (1975). In particular, Klakattawi et al. (2018) recently pointed out that the corresponding regression model can model over- and underdispersed count data. Moreover, they showed that the model is able to adequately fit highly skewed count data with excessive zeros, without the need for introducing zero-inflated or hurdle components, in contrast to other existing methods, e.g., the zero-inflated Conway-Maxwell-Poisson (ZICOM) model (Sellers and Raim, 2016). A further generalization of this approach was introduced by Nekoukhou and Bidram (2015), where the exponentiated discrete Weibull (EDW) distribution is defined.
Apart from the presence of extra-dispersion, extended structures such as longitudinally collected data, where subjects/patients are repeatedly measured over time, and hierarchical structures, originating from hierarchical designs such as multi-centre trials, can also be present. For the GLM framework, the generalized linear mixed model (GLMM), discussed by Engel and Keen (1994), Breslow and Clayton (1993), and Wolfinger and O’Connell (1993), has been suggested, and became a popular framework for taking into account hierarchical data structures. In these models, random effects are introduced to capture the association structure and to some extent dispersion. Molenberghs et al. (2007) extended this approach by introducing the so-called combined modeling (CM) framework, that was mainly developed to encompass both aspects, (1) dispersion and (2) hierarchical/longitudinal structures, simultaneously by adding an extra random effect into the GLMM framework.

In this paper, we examine the existing (univariate) discrete Weibull-based approach of Nakagawa and Osaki (1975), and extend it with random effects to accommodate more complex data structures. This approach assumes that extra dispersion is captured in the pre-specified distribution, and differs from that in Molenberghs et al. (2007) where the extra dispersion is captured by an additional random effect. In addition, various settings (heavy tails, zero-inflation) in combination with dispersion and correlation are examined, and compared with other well-known count models (Appendix A). Conclusions are supported with some characteristics of the model.

The remainder of our paper is organized as follows. In Section 2, two motivating case studies are presented, stemming from patients with epileptic seizures, and historic data on household members from a Belgian town. An overview of the discrete Weibull version of Nakagawa and Osaki (1975) is sketched in Section 3, alongside its extended
version and characteristics. Section 4 is devoted to the analysis of our case studies, where a comparison is made between this approach and other count models (Appendix A). A simulation study is reported in Section 5 to investigate other characteristics of the framework, and concluding remarks are given in Section 6.

2 Case studies

2.1 Epilepsy dataset

The epilepsy dataset comes from a randomized, double-blinded, parallel group multi-center study aimed at comparing placebo with a new anti-epileptic drug (AED), in combination with one or two other AEDs. In total, 45 patients were assigned to the placebo group, and 44 to the active (new) treatment group. Patients were then followed for several weeks – during which the number of epileptic seizures experienced in the last week - were counted, i.e., since the last time the outcome was measured. The main research question is whether or not the new treatment reduces the number of epileptic seizures. A full description of the epilepsy dataset is provided in Faught et al. (2011). Figure 1 shows the individual profiles with corresponding mean and median profiles of the seizure counts for every study week (top), and observed mean and variance of the seizure counts per patient ID (bottom), categorized for both treatment groups. A highly variable longitudinal count data is given with presence of extreme values, zero-inflation, and very few observations available at some of the time-points, especially past week 20.
Figure 1: Epilepsy data (Faught et al., 2011). Subject specific profiles (grey) with corresponding average (solid black) and median (dashed black) profiles of the number of epileptic attacks for every visit (top), and observed mean and variance of the seizure counts per patient ID (bottom), categorized for both treatments.

2.2 Moerzeke dataset

The second dataset comes from a demographic, historical database of Moerzeke, a small village in the center of Flanders (the Dutch speaking part of Belgium) within the province of East Flanders. Information in the database is drawn from church and civil registers, which can be taken as high quality and appropriate for population studies, and includes all individuals who were born, married, or died in Moerzeke.

In this study, a sample of 457 families was taken - by selecting all fathers born between
1750 and 1830, and then forming a family by including also their first born children and children’s mothers. To avoid overlap, children already selected are not included again, either as father or mother of new families. For the group under study, the mean age at death for those who were born and deceased in Moerzeke was 71.9 years for men and 71.7 for women, respectively. The main interest lies in the exploration of different social and/or household characteristics (e.g., gender of first born child) on the (discrete) life expectancy of family members. Figure 2 shows the household profiles with corresponding average and median profiles of the (discrete) life expectancy (top), and observed mean and variance of the (discrete) life expectancy per household ID (bottom), categorized for the gender of the first born child. On the average and median scales, a higher life expectancy of first born male children is observed compared to first born female children.
Figure 2: Moerzeke data. Household specific profiles (grey) with corresponding average (solid black) and median (dashed black) profiles of the (discrete) life expectancy (rescaled) for every household member (top), and observed mean and variance of the (discrete) life expectancy (rescaled) per household ID (bottom), categorized for the gender of the first born child. The indexes F, M, and C refer to father, mother, and first born child, respectively.

3 The Weibull-count approach

Due to the reproductive property of the Gamma distribution, i.e., the sum of two Gamma distributed random variables again follows a Gamma distribution, Winkelmann (1995) pointed out that the Gamma distribution is a useful choice for his two-step approach (Section 1). Unfortunately, this reproductive property does not hold for the Weibull distribution. As an alternative, the discrete approach of Nakagawa
and Osaki (1975), which is here referred as the discrete Weibull (DW) model, can be used instead and gives a simple and adaptable alternative for the Weibull case. In what follows, we will give a general overview of the DW approach of Nakagawa and Osaki (1975).

Let \( Y_i, i = 1, \ldots, n, \) be (type 1) DW distributed (Nakagawa and Osaki, 1975) with parameters \( 0 < q < 1 \) and \( \rho > 0. \) The probability mass function, cumulative distribution function, and hazard function are given by

\[
P(Y_i = y_i) = q^{y_i^\rho} - q^{(y_i+1)^\rho}, \quad F(y_i) = 1 - q^{(y_i+1)^\rho},
\]

\[
h(y_i) = q^{y_i^\rho - (y_i+1)^\rho} - 1,
\]

respectively. Special cases result from this. When \( \rho = 1 \) and \( q = 1 - p, \) the geometric distribution follows. Particularly, when \( \rho = 1 \) and \( q = e^{-\lambda}, \) the discrete exponential (DE) distribution results (Sato et al., 1999), which is overdispersed relative to the standard Poisson distribution (Appendix B). In addition, when \( \rho = 2 \) and \( q = \theta, \) the discrete Rayleigh (DR) distribution of Roy (2004) obtains. If \( \rho \to +\infty, \) the DW approaches a Bernoulli distribution with probability \( q. \) When \( q \) is small, an excessive zero case occurs (Klakattawi et al., 2018).

The mean and variance of the DW are given by

\[
E(Y_i) = \mu = \sum_{n=1}^{+\infty} q^{n^\rho},
\]

\[
\text{Var}(Y_i) = 2 \cdot \sum_{n=1}^{+\infty} n \cdot q^{n^\rho} - \mu - \mu^2.
\]

It can easily be shown that both of these infinite series converge (Appendix C). Based on the integral test, general approximations can be found consisting of incomplete gamma functions, e.g., Englehardt and Li (2011).
To explore the characteristics of the DW model, we compute indexes for dispersion (DI), zero-inflation (ZI) and heavy-tail (HI), which are respectively given by

\[
\text{DI} = \frac{\text{Var}(Y_i)}{E(Y_i)}, \quad \text{ZI} = 1 + \frac{\log P(Y_i=0)}{E(Y_i)}, \quad \text{HT} = \frac{P(Y_i=y_i+1)}{P(Y_i=y_i)} \quad \text{for} \quad y_i \to \infty.
\]

Figure 3: Characteristic indexes related to the Poisson distribution. Dashed, dot-dashed and dotted lines represent the Poisson, DE and DR distribution, respectively.

Note that these indices are defined in relation to the Poisson distribution. Thus, the
dispersion index indicates over-, under- and equidispersion for, respectively, DI > 1, DI < 1 and DI = 1. The zero-inflation index indicates zero-inflation for ZI > 0, zero-deflation for ZI < 0 and no excess of zeros for ZI = 0. At last, heavy-tail index indicates a heavy-tail distribution for HT → 1 when y → ∞. Figure 3 shows that the DW framework is able to model high overdispersed, zero-inflated and heavy-tailed data. The approach also allows the fit of low overdispersed, zero-deflated data, and even some amount of underdispersion.

In a regression framework, Klakattawi et al. (2018) assumed that the response $Y_i$ has a DW distribution, where a subject-specific parameter $q_i$ is related to a $p$-dimensional vector of covariates $x_i$ for $i$th observation through the complementary log-log link function

$$\ln[-\ln(q_i)] = x_i' \cdot \beta$$

$$\Leftrightarrow$$

$$q_i = e^{-e^{x_i' \cdot \beta}}(= e^{-\lambda_i}).$$

Here, $\beta$ represents the associated regression parameter vector, which can directly be interpreted towards the log of the (closed-form) median. This is of particular interest when modelling, for example, highly skewed data, which often occurs in count data. In particular, by subsetting the set of parameters $\beta$ in $\{\beta_0\} \cup \{\beta_l \mid l = 1, \ldots, p\}$, it can easily be shown, thanks to the use of the complementary log-log link function (Klakattawi et al., 2018), that $\{\log[\log(2)] - \beta_0\}/\rho$ is related to the conditional median when all covariates are set to zero, whereas $-\beta_l/\rho$ ($l = 1, \ldots, p$) can be related to the change in the median of the response corresponding to a one unit change of $x_{li}$, keeping all other covariates constant.
In terms of estimation procedures, Klakattawi et al. (2018) and Kulasekera (1994) used maximum likelihood for parameter estimation, while Haselimashhadi et al. (2017) proposed a Bayesian approach for estimating the parameters.

3.1 The extended hierarchical Weibull-count approach

If the discrete data are hierarchically structured, with \( Y_{ij} \) denoting the \( j \)th discrete outcome measured for cluster (subject) \( i \), \( i = 1, \ldots, N \), \( j = 1, \ldots, n_i \), univariate models are often not appropriate to take into account the underlying correlation structure of the data. Therefore, mixed-effects models are often proposed where, in addition to fixed effects, random effects are added to the model to allow for the correlation structure of the data. These approaches have been studied extensively in the GLM framework, e.g., LMM and GLMM (Section 1), while little research has focussed on dispersion models outside of this framework. In our context, where the focus is on the DW approach, a dispersion model, extensions with random effects can simply be achieved as follows:

\[
\ln(\lambda_{ij}) = \mathbf{x}_{ij}' \cdot \beta + \mathbf{z}_{ij}' \cdot \mathbf{b}_i,
\]

where \( \mathbf{z}_{ij} \) represents a \( q \)-dimensional vector of known covariate values corresponding to the \( q \)-dimensional random effects vector \( \mathbf{b}_i \) following a multivariate normal distribution with mean vector \( \mathbf{0} \) and variance-covariance matrix \( D \).

In the following, we will analyze the epilepsy and Moerzeke datasets, introduced in Section 2. Maximum likelihood principles are used to obtain valid parameter estimates. The SAS procedure NLMIXED is used for the computations (Appendix D).
4 Analysis of case studies

4.1 Epilepsy dataset

The epilepsy data of Section 2 will be analyzed with the DW and its nested DE model (Section 3), and compared with some conventional models from Appendix A, i.e., the classical Poisson log-linear, negative binomial (NB), Conway-Maxwell-Poisson (COM) and DP models. Previous work on this dataset was reported by Molenberghs and Verbeke (2005) and Molenberghs et al. (2007) in the context of generalized estimating equations (Liang and Zeger, 1986) and the CM framework, respectively.

Let $Y_{ij}$ be the number of epileptic seizures that patient $i$ experiences during week $j$ of the follow-up period, and let $t_{ij}$ be the time-point at which outcome $Y_{ij}$ has been measured, i.e., $t_{ij} = 1, 2, \ldots$ until at most 27. The following specific choice is made for the link function:

$$
\ln(\lambda_{ij}) = \beta_0 + b_i + \beta'_0 \cdot T_i + (\beta_1 + \beta'_1 \cdot T_i) \cdot t_{ij},
$$

where $T_i = 1$ if patient $i$ receives the treatment, and 0 for placebo. Here, $\beta'_0$ and $\beta'_1$ represent differences between treatment and placebo in terms of intercept and slope, respectively. The random intercept $b_i$ is assumed to be normally distributed with mean 0 and variance $\sigma^2$, reflecting the between-patient variability within the data. Maximum likelihood estimates and corresponding standard errors of the parameters are reported in Table 1 (for the univariate case, i.e., without the subject random effect) and Table 2 (for the clustered case, i.e., with the subject random effect).
Table 1: Epilepsy dataset. Parameter estimates and standard errors for the regression coefficients in (1) the Poisson (P) model, (2) discrete exponential (DE) approach, (3) negative binomial (NB) model, (4) Conway-Maxwell-Poisson (COM) model, (5) double Poisson (DP) model, and (6) the discrete Weibull (DW) model.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>( P)</th>
<th>( DE)</th>
<th>( NB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept placebo</td>
<td>( \beta_0 )</td>
<td>1.2662 (0.0424)</td>
<td>1.2601 (0.0864)</td>
<td>1.2594 (0.1119)</td>
</tr>
<tr>
<td>Difference in intercepts</td>
<td>( \beta'_0 )</td>
<td>0.1869 (0.0571)</td>
<td>0.2115 (0.1202)</td>
<td>0.2156 (0.1564)</td>
</tr>
<tr>
<td>Slope placebo</td>
<td>( \beta_1 )</td>
<td>-0.0134 (0.0043)</td>
<td>-0.0126 (0.0086)</td>
<td>-0.0126 (0.0111)</td>
</tr>
<tr>
<td>Difference in slopes</td>
<td>( \beta'_1 )</td>
<td>-0.0195 (0.0058)</td>
<td>-0.0222 (0.0116)</td>
<td>-0.0227 (0.0150)</td>
</tr>
<tr>
<td>Ratio of slopes</td>
<td>( 1 + \frac{\beta'_1}{\beta_1} )</td>
<td>2.4576 (0.8480)</td>
<td>2.7586 (1.9721)</td>
<td>2.8081 (2.6066)</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>1.8961 (0.0918)</td>
</tr>
<tr>
<td>( \tau )</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>( \phi )</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>( \rho )</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>( -2 \text{ loglik} )</td>
<td>11590.0</td>
<td>6502.5</td>
<td>6326.1</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>11598.0</td>
<td>6510.5</td>
<td>6336.1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>( \text{COM} )</th>
<th>( \text{DP} )</th>
<th>( \text{DW} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept placebo</td>
<td>( \beta_0 )</td>
<td>-0.5054 (0.0189)</td>
<td>1.2662 (0.1054)</td>
<td>0.7341 (0.1002)</td>
</tr>
<tr>
<td>Difference in intercepts</td>
<td>( \beta'_0 )</td>
<td>0.0131 (0.0144)</td>
<td>0.1869 (0.1421)</td>
<td>0.0936 (0.1307)</td>
</tr>
<tr>
<td>Slope placebo</td>
<td>( \beta_1 )</td>
<td>-0.0011 (0.0012)</td>
<td>-0.0134 (0.0108)</td>
<td>-0.0174 (0.0095)</td>
</tr>
<tr>
<td>Difference in slopes</td>
<td>( \beta'_1 )</td>
<td>-0.0017 (0.0017)</td>
<td>-0.0195 (0.0144)</td>
<td>-0.0143 (0.0127)</td>
</tr>
<tr>
<td>Ratio of slopes</td>
<td>( 1 + \frac{\beta'_1}{\beta_1} )</td>
<td>2.5663 (3.1297)</td>
<td>2.4576 (2.1093)</td>
<td>1.8189 (1.1027)</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>( \tau )</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>( \phi )</td>
<td>--</td>
<td>0.1616 (0.0061)</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>( \rho )</td>
<td>--</td>
<td>--</td>
<td>0.7383 (0.0172)</td>
<td>--</td>
</tr>
<tr>
<td>( -2 \text{ loglik} )</td>
<td>6256.2</td>
<td>6815.6</td>
<td>6291.3</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>6266.2</td>
<td>6825.6</td>
<td>6301.3</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Epilepsy dataset. Parameter estimates and standard errors for the regression coefficients in (1) the Poisson-normal (PN) model, (2) discrete exponential-normal (DEN) approach, (3) combined (CM) model, (4) Conway-Maxwell-Poisson-normal (COMN) model, (5) double Poisson-normal (DPN) model, and (6) the discrete Weibull-normal (DWN) model.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>PN</th>
<th>DEN</th>
<th>CM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Est. (s.e.)</td>
<td>Est. (s.e.)</td>
<td>Est. (s.e.)</td>
</tr>
<tr>
<td>Intercept placebo</td>
<td>$\beta_0$</td>
<td>0.8177 (0.1677)</td>
<td>0.9443 (0.1843)</td>
<td>0.9112 (0.1755)</td>
</tr>
<tr>
<td>Difference in intercepts</td>
<td>$\beta_0'$</td>
<td>-0.1705 (0.2387)</td>
<td>-0.2670 (0.2620)</td>
<td>-0.2556 (0.2500)</td>
</tr>
<tr>
<td>Slope placebo</td>
<td>$\beta_1$</td>
<td>-0.0143 (0.0044)</td>
<td>-0.0271 (0.0101)</td>
<td>-0.0248 (0.0077)</td>
</tr>
<tr>
<td>Difference in slopes</td>
<td>$\beta_1'$</td>
<td>0.0023 (0.0062)</td>
<td>0.0145 (0.0140)</td>
<td>0.0130 (0.0107)</td>
</tr>
<tr>
<td>Ratio of slopes</td>
<td>$1 + \frac{\beta_1'}{\beta_1}$</td>
<td>0.8398 (0.3979)</td>
<td>0.4663 (0.3953)</td>
<td>0.4751 (0.3345)</td>
</tr>
<tr>
<td>Std. dev. random effect</td>
<td>$\sigma$</td>
<td>1.0755 (0.0857)</td>
<td>1.0436 (0.0888)</td>
<td>1.0626 (0.0871)</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>--</td>
<td>--</td>
<td>0.4059 (0.0348)</td>
</tr>
<tr>
<td></td>
<td>$\tau$</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>$-2$ loglik</td>
<td></td>
<td>6271.9</td>
<td>5543.9</td>
<td>5417.0</td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td>6281.9</td>
<td>5553.9</td>
<td>5429.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>COMN</th>
<th>DPN</th>
<th>DWN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Est. (s.e.)</td>
<td>Est. (s.e.)</td>
<td>Est. (s.e.)</td>
</tr>
<tr>
<td>Intercept placebo</td>
<td>$\beta_0$</td>
<td>-0.2384 (0.0779)</td>
<td>0.8314 (0.1721)</td>
<td>1.4319 (0.2183)</td>
</tr>
<tr>
<td>Difference in intercepts</td>
<td>$\beta_0'$</td>
<td>-0.0947 (0.1042)</td>
<td>-0.1582 (0.2451)</td>
<td>-0.2970 (0.3005)</td>
</tr>
<tr>
<td>Slope placebo</td>
<td>$\beta_1$</td>
<td>-0.0040 (0.0023)</td>
<td>-0.0146 (0.0067)</td>
<td>-0.0297 (0.0098)</td>
</tr>
<tr>
<td>Difference in slopes</td>
<td>$\beta_1'$</td>
<td>0.0005 (0.0032)</td>
<td>0.0018 (0.0093)</td>
<td>0.0180 (0.0135)</td>
</tr>
<tr>
<td>Ratio of slopes</td>
<td>$1 + \frac{\beta_1'}{\beta_1}$</td>
<td>0.8646 (0.7451)</td>
<td>0.8778 (0.5980)</td>
<td>0.3947 (0.3382)</td>
</tr>
<tr>
<td>Std. dev. random effect</td>
<td>$\sigma$</td>
<td>0.4475 (0.0433)</td>
<td>1.0458 (0.0875)</td>
<td>1.2658 (0.1063)</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>$\tau$</td>
<td>0.1563 (0.0196)</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>$\phi$</td>
<td>--</td>
<td>0.4355 (0.0169)</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>--</td>
<td>--</td>
<td>1.3074 (0.0340)</td>
</tr>
<tr>
<td>$-2$ loglik</td>
<td></td>
<td>5473.8</td>
<td>5652.2</td>
<td>5451.1</td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td>5485.8</td>
<td>5664.2</td>
<td>5463.1</td>
</tr>
</tbody>
</table>
In the univariate case, i.e., where clustering is ignored (Table 1), we clearly observe very large improvements in the DE, NB, DP and DW models, in terms of the likelihood, relative to the classical Poisson model. This, of course, is to be expected since the Poisson model assumes equidispersion while the parameters $\alpha$, $\phi$ and $\rho$ (see Appendix A for details) provide significant evidence of overdispersion. Furthermore, when a comparison is made between the DW and the conventional models, e.g., NB and DP, w.r.t. the Poisson model, we could consider the DW model as the better one in terms of log-likelihood. Indeed, similar to the NB approach (Figure 7), the DW model is able to capture highly overdispersed, zero-inflated and heavy-tailed data (Figure 3), characteristics that are definitely present within the epilepsy dataset. Bar charts of the fitted univariate models are given in Figure 4.

Furthermore, we should mention that ‘illegal’ estimates were obtained for the COM model, implying that no valid conclusions can be made from it. Indeed, when looking at the fitted dispersion parameter $\tau$, a negative estimate ($-0.1188$) is observed which is outside the parameter space (Appendix A3). This, of course, can easily be explained by the fact that the COM distribution limits itself in flexibility towards underdispersed data with narrow flexibility to zero-inflation (Figure 8).

For the clustered case, i.e., where a subject specific random intercept is added to account for correlation (Table 2), we find that the DWN is considerably better in terms of likelihood. Moreover, point and precision estimates of such key parameters as the slope difference and the slope ratio are strongly affected when a random effect is added to the models. This remark was also made by Molenberghs et al. (2010), who noted an impact on hypothesis testing. Surprisingly, a valid interpretation on the extended COM approach can now be given, while this was not possible in the univariate case. To
Figure 4: Bar charts of fitted univariate models
explain this phenomenon, attention should be directed towards the limited flexibility of COM in terms of overdispersion and the multiplicity effect of the random effects. In particular, a limited number of highly overdispersed regions can be modelled with the COM approach (Figure 8). By adding a random effect to the model, extra flexibility has been given towards capturing overdispersed regions. Indeed, since random effects are mainly used to capture the underlying correlation structure of the data, they are also able to seize a certain amount of dispersion. Therefore, more flexibility has been gained with the inclusion of random effects towards the modelling of overdispersed data. In addition, a much lower parameter estimate for $\sigma$ was obtained for the COMN case, compared to all other models. This directly results from the main disadvantage of the COM regression model, i.e., its location parameter does not correspond to the expectation, which complicates the interpretation of regression models towards the mean specified using this distribution (Sellers and Shmueli, 2010).

Even though the CM model is a more viable candidate in terms of likelihood (related to the Poisson model), one should be aware of the restricted mean scale interpretation in this framework, especially when dealing with skewed data. In this setting, right skewed data (Figure 1) is observed, making the inferences less attractive from an interpretational point of view (similar to the DPN approach). The DW model avoids this problem by allowing inferences directly on the median scale (Section 3), making the approach more interesting here.

Finally, we expand our analysis with random slopes in the DWN model, i.e., considering two random effects instead of a single to reflect the between- and within-patient variability of the data. The predictor function becomes:

$$\ln(\lambda_{ij}) = \beta_0 + b_{1i} + \beta_0' \cdot T_i + (\beta_1 + \beta_1' \cdot T_i + b_{2i}) \cdot t_{ij},$$
where the random effects vector $\mathbf{b}_i = (b_{1i}, b_{2i})'$ is assumed to be multivariate normally distributed with mean vector $\mathbf{0}$ and variance-covariance matrix

$$D = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$ 

A comparison with the random-intercept model will be made in two ways, i.e., (1) a random-slopes model with uncorrelated random effects ($\sigma_{12} = 0$; IND) and (2) a random-slopes model with correlated random effects ($\sigma_{12} \neq 0$; UN). Maximum likelihood estimates and corresponding standard errors of the parameters are reported in Table 3.

Table 3: Epilepsy dataset. Parameter estimates and standard errors for the regression coefficients in the discrete Weibull-normal (DWN) model with (1) random-intercept, (2) random-slope with uncorrelated random effects (IND), and (3) random-slope with correlated random effects (UN).

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>Random-intercept</th>
<th>Random-slope (IND)</th>
<th>Random-slope (UN)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Est. (s.e.)</td>
<td>Est. (s.e.)</td>
<td>Est. (s.e.)</td>
</tr>
<tr>
<td>Intercept placebo</td>
<td>$\beta_0$</td>
<td>1.4319 (0.2183)</td>
<td>1.4973 (0.2183)</td>
<td>1.4947 (0.2287)</td>
</tr>
<tr>
<td>Difference in intercepts</td>
<td>$\beta_0'$</td>
<td>-0.2970 (0.3005)</td>
<td>-0.2909 (0.2996)</td>
<td>-0.2984 (0.3150)</td>
</tr>
<tr>
<td>Slope placebo</td>
<td>$\beta_1$</td>
<td>-0.0297 (0.0098)</td>
<td>-0.0339 (0.0120)</td>
<td>-0.0327 (0.0126)</td>
</tr>
<tr>
<td>Difference in slopes</td>
<td>$\beta_1'$</td>
<td>0.0180 (0.0135)</td>
<td>0.0169 (0.0168)</td>
<td>0.0167 (0.0176)</td>
</tr>
<tr>
<td>Ratio of slopes</td>
<td>$1 + \frac{\beta_1}{\beta_0}$</td>
<td>0.3947 (0.3382)</td>
<td>0.5016 (0.3920)</td>
<td>0.4884 (0.4219)</td>
</tr>
<tr>
<td>Std. dev. random-intercept</td>
<td>$\sigma_1$</td>
<td>1.2658 (0.1063)</td>
<td>1.2553 (0.1114)</td>
<td>1.3333 (0.1302)</td>
</tr>
<tr>
<td>Std. dev. random-slope</td>
<td>$\sigma_2$</td>
<td>-- (--)</td>
<td>0.0417 (0.0092)</td>
<td>0.0474 (0.0099)</td>
</tr>
<tr>
<td>Cov. between random-effects</td>
<td>$\sigma_{12}$</td>
<td>-- (--)</td>
<td>-- (--)</td>
<td>-0.0177 (0.0142)</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>1.3074 (0.0340)</td>
<td>1.3393 (0.0362)</td>
<td>1.3463 (0.0366)</td>
</tr>
</tbody>
</table>

-2 loglik                     |      | 5451.1               | 5439.6             | 5437.7            |
AIC                            |      | 5463.1               | 5453.6             | 5453.7            |

A significant improvement in likelihood is observed when adding a random-slope to the
model (likelihood ratio test $p = 0.0007$). However, there are no qualitative changes in the results of hypothesis testing for the main effects of interest. Furthermore, by comparing the independent random effects (IND) with correlated random effects (UN), no significant improvements were obtained (likelihood ratio test $p = 0.1692$). This extension at the same time illustrates the ease with which more than one random effect can be included.

4.2 Moerzeke dataset

While previous work on the Moerzeke data was provided by Matthijs et al. (2002) for the examination of historical mortality in terms sociological and biological components, there has been no consideration of dispersion aspects. To this end, the DW and its nested DE models are considered in the analysis of the Moerzeke dataset (Section 2) and compared with the count models from Appendix A.

Let $Y_{ij}$ represent the (discrete) life expectancy of the mother, father, and first born child ($j = 1, 2, 3$) in household $i = 1, \ldots, 457$, and write $G_i$ as the gender effect of first born child in household $i$. Here, we assume the following log-linear model:

$$
\ln(\lambda_{ij}) = \beta_0 \cdot I_{C_{ij}} + \beta'_0 \cdot I_{M_{ij}} + \beta''_0 \cdot I_{F_{ij}} + b_i + (\beta_1 \cdot I_{C_{ij}} + \beta'_1 \cdot I_{M_{ij}} + \beta''_1 \cdot I_{F_{ij}}) \cdot G_i,
$$

with $I_{M_{ij}}$, $I_{F_{ij}}$ and $I_{C_{ij}}$ representing binary indicators for mother, father and first born child, respectively, i.e., $I_{M_{ij}} = 1$ if household member $j$ in household $i$ is the mother, 0 otherwise; $I_{F_{ij}} = 1$ if household member $j$ in household $i$ is the father, 0 otherwise; and $I_{C_{ij}} = 1$ if household member $j$ in household $i$ is the first born child, 0 otherwise. Here, different baseline and gender of first born child effects per household members are assumed. Again, a random intercept $b_i$ is used to capture between-
Martial Luyts et al.

household variability, which here is assumed normally distributed with mean 0 and variance $\sigma^2$. Maximum likelihood estimates and corresponding standard errors of the parameters are reported in Table 4 (for the univariate case without the random effect) and Table 5 (for the clustered case, including with random effect).

In the univariate case (Table 4), the COM, DP and DW models significantly improved the model fit, compared to the classical Poisson model, while, in terms of likelihood, a worse fit is observed for the DE case. Indeed, when considering the dispersion parameters ($\tau, \phi$ and $\rho$), we observe the clear presence of underdispersion within the data. While the COM, DP and DW models are able to capture this phenomenon (Figures 3, 7, and 8), this is not the case for the DE (Appendix B) and Poisson models. Therefore, it is fair to say that the DE model is completely wrong, not just in terms of underdispersion, but also in the fact that it fails to capture the unimodal shape, as expected from a geometric distribution. The underdispersion result can be explained by the fact that Moerzeke has characteristics of a geographically isolated area as it is almost completely surrounded by a meander in the river Scheldt and by the river Durme. This was an important geographical limitation within the time bracket at which data were collected, and led to more genetic homogeneity than in the typical town. We observe that the DW model indicates the best fit, relative to the COM and DP models, in terms of likelihood w.r.t. the Poisson model. A possible reason for this result is the presence of left-skewed discrete time-to-event data, which can flexibly be modelled with the DW approach due to the underlying Weibull connection. Bar charts of the fitted univariate models are given in Figure 5.

In the clustered case (Table 5), noteworthy results were obtained for the estimated variance component $\sigma^2$. In all clustered models, the estimated component is very close
Table 4: Moerzeke dataset. Parameter estimates and standard errors for the regression coefficients in (1) the Poisson (P) model, (2) the discrete exponential (DE) approach, (3) Conway-Maxwell-Poisson (COM) model, (4) double Poisson (DP) model, and (5) the discrete Weibull (DW) model.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>Est. (s.e.)</th>
<th>DE</th>
<th>Est. (s.e.)</th>
<th>Est. (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept first born child</td>
<td>$\beta_0$</td>
<td>1.7068 (0.0288)</td>
<td>1.7068 (0.0735)</td>
<td>3.0527 (0.1294)</td>
<td></td>
</tr>
<tr>
<td>Intercept father</td>
<td>$\beta_0'$</td>
<td>1.8473 (0.0268)</td>
<td>1.8473 (0.0727)</td>
<td>3.2891 (0.1373)</td>
<td></td>
</tr>
<tr>
<td>Intercept mother</td>
<td>$\beta_0''$</td>
<td>1.8847 (0.0263)</td>
<td>1.8847 (0.0725)</td>
<td>3.3522 (0.1395)</td>
<td></td>
</tr>
<tr>
<td>Gender effect on first born child</td>
<td>$\beta_1$</td>
<td>0.1009 (0.0390)</td>
<td>0.1009 (0.1014)</td>
<td>0.1697 (0.0509)</td>
<td></td>
</tr>
<tr>
<td>Gender effect on father</td>
<td>$\beta_1'$</td>
<td>0.0187 (0.0370)</td>
<td>0.0187 (0.1007)</td>
<td>0.03163 (0.0481)</td>
<td></td>
</tr>
<tr>
<td>Gender effect on mother</td>
<td>$\beta_1''$</td>
<td>0.0145 (0.0364)</td>
<td>0.0145 (0.1005)</td>
<td>0.02474 (0.0473)</td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>1.7484 (0.0690)</td>
<td></td>
</tr>
<tr>
<td>$\phi$</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

-2 loglik 5834.3 7985.1 5669.3
AIC 5846.3 7997.1 5683.3

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>Est. (s.e.)</th>
<th>DP</th>
<th>Est. (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept first born child</td>
<td>$\beta_0$</td>
<td>1.7068 (0.0225)</td>
<td>8.9228 (0.2301)</td>
<td></td>
</tr>
<tr>
<td>Intercept father</td>
<td>$\beta_0'$</td>
<td>1.8473 (0.0210)</td>
<td>9.0796 (0.2293)</td>
<td></td>
</tr>
<tr>
<td>Intercept mother</td>
<td>$\beta_0''$</td>
<td>1.8847 (0.02061)</td>
<td>9.1660 (0.2301)</td>
<td></td>
</tr>
<tr>
<td>Gender effect on first born child</td>
<td>$\beta_1$</td>
<td>0.1009 (0.03048)</td>
<td>0.1699 (0.0957)</td>
<td></td>
</tr>
<tr>
<td>Gender effect on father</td>
<td>$\beta_1'$</td>
<td>0.0187 (0.0290)</td>
<td>0.0831 (0.0955)</td>
<td></td>
</tr>
<tr>
<td>Gender effect on mother</td>
<td>$\beta_1''$</td>
<td>0.0145 (0.0285)</td>
<td>0.0350 (0.0954)</td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>$\phi$</td>
<td>--</td>
<td>1.6333 (0.0624)</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>$\rho$</td>
<td>--</td>
<td>--</td>
<td>4.5377 (0.1055)</td>
<td>--</td>
</tr>
</tbody>
</table>

-2 loglik 5693.3 5512.3
AIC 5707.3 5526.3
Table 5: *Moerzeke dataset. Parameter estimates and standard errors for the regression coefficients in (1) the Poisson-normal (PN) model, (2) the discrete exponential-normal (DEN) approach, (3) Conway-Maxwell-Poisson-normal (COMN) model, (4) double Poisson-normal (DPN) model, and (5) the discrete Weibull-normal (DWN) model.*

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>PN</th>
<th>DEN</th>
<th>COMN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Est. (s.e.)</td>
<td>Est. (s.e.)</td>
<td>Est. (s.e.)</td>
</tr>
<tr>
<td>Intercept first born child</td>
<td>$\beta_0$</td>
<td>1.7068 (0.0288)</td>
<td>1.7068 (0.0735)</td>
<td>3.0529 (0.1294)</td>
</tr>
<tr>
<td>Intercept father</td>
<td>$\beta_0$</td>
<td>1.8473 (0.0268)</td>
<td>1.8472 (0.0727)</td>
<td>3.2895 (0.1373)</td>
</tr>
<tr>
<td>Intercept mother</td>
<td>$\beta_0$</td>
<td>1.8847 (0.0263)</td>
<td>1.8847 (0.0727)</td>
<td>3.3527 (0.1395)</td>
</tr>
<tr>
<td>Gender effect on first born child</td>
<td>$\beta_1$</td>
<td>0.1009 (0.0390)</td>
<td>0.1009 (0.1014)</td>
<td>0.1698 (0.0509)</td>
</tr>
<tr>
<td>Gender effect on father</td>
<td>$\beta_1$</td>
<td>0.0187 (0.0370)</td>
<td>0.0187 (0.1007)</td>
<td>0.0317 (0.0481)</td>
</tr>
<tr>
<td>Gender effect on mother</td>
<td>$\beta_1$</td>
<td>0.0145 (0.0364)</td>
<td>0.0145 (0.1005)</td>
<td>0.02453 (0.0473)</td>
</tr>
<tr>
<td>Std. dev. random effect</td>
<td>$\sigma$</td>
<td>$1.16 \times 10^{-4}$ (0.0119)</td>
<td>$7.72 \times 10^{-4}$ (0.1039)</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau$</td>
<td>---</td>
<td>1.7486 (0.0690)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\phi$</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\rho$</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

-2 loglik: PN 5834.3, DEN 7985.1, COMN 5669.3, DPN 5848.3, DWN 7999.1, AIC 5685.3

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>DPN</th>
<th>DWN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Est. (s.e.)</td>
<td>Est. (s.e.)</td>
</tr>
<tr>
<td>Intercept first born child</td>
<td>$\beta_0$</td>
<td>1.7068 (0.0225)</td>
<td>8.9228 (0.2301)</td>
</tr>
<tr>
<td>Intercept father</td>
<td>$\beta_0$</td>
<td>1.8473 (0.0210)</td>
<td>9.0795 (0.2293)</td>
</tr>
<tr>
<td>Intercept mother</td>
<td>$\beta_0$</td>
<td>1.8846 (0.0206)</td>
<td>9.1660 (0.2301)</td>
</tr>
<tr>
<td>Gender effect on first born child</td>
<td>$\beta_1$</td>
<td>0.1010 (0.0305)</td>
<td>0.1699 (0.0957)</td>
</tr>
<tr>
<td>Gender effect on father</td>
<td>$\beta_1$</td>
<td>0.0187 (0.0290)</td>
<td>0.0831 (0.0955)</td>
</tr>
<tr>
<td>Gender effect on mother</td>
<td>$\beta_1$</td>
<td>0.0145 (0.0285)</td>
<td>0.0350 (0.0954)</td>
</tr>
<tr>
<td>Std. dev. random effect</td>
<td>$\sigma$</td>
<td>$1.85 \times 10^{-4}$ (0.0293)</td>
<td>$2.33 \times 10^{-4}$ (0.0420)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau$</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\phi$</td>
<td>1.6333 (0.0624)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\rho$</td>
<td>---</td>
</tr>
</tbody>
</table>

-2 loglik: DPN 5693.3, DWN 5512.3, AIC 5709.3, 5528.3
A Weibull-count approach for under- and overdispersed longitudinal/clustered data

(a) Poisson model

(b) Conway-Maxwell-Poisson model

(c) Double Poisson model

(d) Discrete exponential model

(e) Discrete Weibull model

Figure 5: Bar charts of fitted univariate models
to 0, leaving the other parameter estimates unchanged relative to the univariate cases. This phenomenon, while strange at first sight, is reasonably well understood in the literature. More specifically, partial-marginalization is used here, in agreement with Molenberghs et al. (2010), where adaptive Gaussian quadrature principles are used to approximate the marginal likelihood obtained from integrating over the normal random effects. This automatically adopts a hierarchical perspective, implying the restriction that no negative estimates of $\sigma^2$ can be achieved, even though this could be present for several reasons (e.g., negative intra-class correlation, underdispersion, etc.). Molenberghs and Verbeke (2011) and Verbeke and Molenberghs (2003), for example, discussed this phenomenon in the context of linear mixed models. Pryseley et al. (2011) extended this discussion to non-Gaussian outcomes, while Oliveira et al. (2017) illustrated how such negative variance components play a natural role in modeling both the correlation between repeated measures on the same experimental unit and over- or underdispersion from a CM perspective.

To conclude, we should mention that, even though the DW model fits the data quite well in the context of underdispersed data for the univariate case, there is still scope for further research in the context of underdispersed clustered data. Even though it is not our scope to fully encounter this problem here, boundary issues are suggested for the variance component. Also note that the random-effects variability is very different between the epilepsy and Moerzeke studies, underscoring that a large range of situations can be handled. Of course, this does not preclude further research towards underdispersion.
5 Simulation study

To further explore the DW approach with dispersed count data, a simulation study is conducted to examine the deviance surface under different dispersion situations. This highlights some other characteristics of the model such as the orthogonality and computational ease of estimating the parameters \((q, \rho)\).

Figure 6 presents contour plots of the deviance surfaces for five different simulated DW data of size 1000, with expectation fixed at 1 and dispersion indices at 0.25 (very strong underdispersion), 0.5 (strong underdispersion), 1 (equi-dispersion), 5 (strong overdispersion) and 10 (very strong overdispersion). As a result, the figure indicates high correlation between the parameters, which is related to the dispersion index. More specifically, a decreasing trend in the correlation seems to correspond with an increasing dispersion index. Based on the deviance surface, computational ease is combined with the ability to perform asymptotic (normally based) inferences in the regions with high dispersion, i.e., \( DI \rightarrow \infty \).
Figure 6: Deviance surfaces for discrete Weibull model fitted to five simulated data with expectation 1 and dispersion 0.25, 0.5, 1, 5 and 10. Dotted lines are the maximum likelihood estimates, and white points are the parameters used in the simulation.
6 Concluding remarks

Starting from an existing univariate framework, we have proposed an extended version that can handle both under- and overdispersed, and hierarchical data structures. In both case studies, we showed that the model fits the data well, for both under- and overdispersed situations. More specifically, the approach used is able to flexibly model highly overdispersed, zero-inflated, heavy-tailed and correlated data, similar to the CM approach. In addition, the approach is capable of modeling some low overdispersed regions with zero-deflation (e.g., the DR approach for small values of $q$) and even underdispersed data, regions that cannot be captured within the CM framework. Due to the presence of a closed-form median expression, interpretations of the parameters can directly be related to the median profile, which is of particular interest when modelling skewed data. At last, orthogonality properties are examined through a simulation study. The resulting outcome indicates the presence of correlation between parameters, related to the dispersion index.

Acknowledgements

Financial support from the IAP research network #P7/06 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged. This work was partially supported by CNPq, a Brazilian science funding agency. The research leading to these results has also received funding from KU Leuven GOA project: “New approaches to the social dynamics of long term fertility change”. We thank Mr R. Bijl, a Flemish genealogist and member of the V.V.F.-Dendermonde (“Vlaamse Vereniging voor Familiekunde”, Dendermonde), for granting us access to his demographic data on Moerzeke.
References


A Weibull-count approach for handling under- and overdispersed longitudinal/clustered data structures


Supplementary Materials

A General overview of over- and/or underdispersed count models

Due to the restricted mean-variance relationship of the Poisson log-linear GLM models (i.e., $E(Y_i) = Var(Y_i) \equiv \lambda_i$, where $Y_i \in \mathbb{N}$, $i = 1, \ldots, n$, is Poisson distributed with parameter $\lambda_i \in \mathbb{R}_{>0}$, the set of positive real numbers, and with $\ln(\lambda_i) = x_i' \cdot \beta$, where $x_i$ and $\beta$ denote a $p$-dimensional vector of covariates for $i$th observation and the associated parameter vector, respectively), various extensions have been proposed in the literature (Breslow, 1984; Lawless, 1987; Hinde and Demétrio, 1998) - and in what follows, some popular models will be examined in more detail.
1 Quasi-Poisson model

A straightforward modification, in the context of the exponential family, is to allow the dispersion (scale) parameter, denoted by \( \delta \), to not be restricted to 1. This leads us to \( \text{Var}(Y_i) = \delta \cdot E(Y_i) \), where \( \delta > 1 \) and \( \delta < 1 \) indicates over- and underdispersion, respectively. This results in the so-called quasi-Poisson model (Wedderburn, 1974), where the point estimates of \( \beta \) are identical to those of the standard Poisson model, but standard errors are scaled by \( \sqrt{\delta} \) resulting in possible differences in inferences on covariates compared to the standard Poisson model.

2 Negative binomial model

Another elegant way to provide flexibility is through a two-stage model. A popular approach in this context is to assume that \( Y_i \mid \lambda_i \sim \text{Poi}(\lambda_i) \) and that the parameter \( \lambda_i \) is itself a random variable with mean \( \mu_i \) and variance \( \sigma_i^2 \). By using standard results on iterated expectations we have:

\[
\text{E}(Y_i) = \mu_i, \quad \text{Var}(Y_i) = \mu_i + \sigma_i^2.
\]

A popular specific distributional choice is \( \lambda_i \sim \text{Gamma}(\alpha, \alpha^{-1}) \), for reasons of identifiability (Duchateau and Janssen, 2007), leading to the negative binomial (NB) model. Choosing the Gamma distribution has the advantage of (1) satisfying the mean’s scale for count outcomes and (2) obtaining closed forms for the marginal mean and variance, and even for the entire marginal distribution (Molenberghs et al., 2007). The corresponding (marginal) probability mass function, mean and variance of the
model are equal to

\[ P(Y_i = y_i \mid x_i) = \frac{1}{Z(\lambda_i, \tau)} \cdot \frac{\lambda_i^y}{(y_i!)^\tau} \cdot \frac{1}{\sum_{n=0}^{+\infty} \frac{\lambda_i^n}{(n!)^\tau}}, \]

with \( \lambda_i = e^{x_i' \beta} \), 

\[ E(Y_i) = \lambda_i, \quad \text{Var}(Y_i) = \lambda_i + \alpha \cdot \lambda_i^2, \]

respectively (Lawless, 1987; Cameron and Trivedi, 1986). We should note that, from a hierarchical/conditional viewpoint, only overdispersion can be examined (since for a valid Gamma distribution \( \alpha > 0 \)).

### 3 Conway-Maxwell-Poisson model

The Conway-Maxwell-Poisson (COM) model, first introduced by Conway and Maxwell (1962), is suitable for analysing count data that exhibit either over- or underdispersion. Even though its existence has been known for several decades, most research on this model was done during the last decade. Shmueli et al. (2005), for example, investigated the statistical properties of the COM distribution. While in a Bayesian context, Kadane et al. (2006) developed the conjugate distributions for the parameters of the COM distribution. The probability mass function of the model can be expressed as

\[ P(Y_i = y_i \mid x_i) = \frac{1}{Z(\lambda_i, \tau)} \cdot \frac{\lambda_i^y}{(y_i!)^\tau} \cdot \frac{1}{\sum_{n=0}^{+\infty} \frac{\lambda_i^n}{(n!)^\tau}}, \]

The domain of admissible parameters for which the probability mass function above defines a probability distribution is \((\lambda_i, \tau) > 0\), and \(0 < \lambda_i < 1, \tau = 0\). Some well-known discrete data models result from this. When \( \tau \) equals 1, it reduces to the standard Poisson model. When \( \tau \to +\infty \), the COM model approaches a Bernoulli
model with success parameter \( \pi_i = \frac{\lambda_i}{1+\lambda_i} \). While if \( \tau = 0 \) and \( \lambda_i < 1 \), the geometric model with success probability \( 1-\lambda_i \) is obtained. In terms of dispersion, specific focus is put on the mean and variance functions and the nature of the different dispersion regions can be found in Appendix E2.

The mean and variance are given by
\[
E(Y_i) = \lambda_i \cdot \frac{\partial \log Z(\lambda_i, \tau)}{\partial \lambda_i} \approx \lambda_i^{1/\tau} - \frac{\tau-1}{2\tau}, \quad \text{Var}(Y_i) = \frac{\partial E(Y_i)}{\partial \log \lambda_i} \approx \frac{1}{\tau} \cdot \lambda_i^{1/\tau}.
\]

### 4 Double Poisson model

The double Poisson (DP) model, based on the double-exponential family of Efron (1986), has hardly been investigated and applied since its first introduction three decades ago. Winkelmann (2008) and Hilbe (2011) indicated that the normalizing constant is the bottleneck in applying the DP by showing that fitted models with its normalizing constant approximated by Efron’s original method are not exact. For these and other reasons, different approximations have been proposed in the literature. A full discussion can be found in Zou et al. (2013).

The probability mass function of the DP model can be written as
\[
P(Y_i = y_i \mid x_i) = K(\lambda_i, \phi) \cdot \phi^{y_i/2} \cdot e^{-\phi \lambda_i} \cdot \left( \frac{e \cdot \lambda_i}{y_i} \right)^y \cdot \left( 1 + \frac{1-\phi}{12 \cdot \phi \lambda_i} \right) \cdot \left( 1 + \frac{1-\phi}{\phi \lambda_i} \right),
\]
\[
\lambda_i = e^{x_i' \beta}, \quad \frac{1}{K(\lambda_i, \phi)} \approx 1 + \frac{1-\phi}{12 \cdot \phi \lambda_i} \cdot \left( 1 + \frac{1-\phi}{\phi \lambda_i} \right),
\]
where \( K(\lambda_i, \phi) \) is the normalizing constant that is often close to 1. The corresponding mean and variance can be estimated by
\[
E(Y_i) \approx \lambda_i, \quad \text{Var}(Y_i) \approx \frac{\lambda_i}{\phi}.
\]
Thus, the DP model allows for both overdispersion ($\phi < 1$) and underdispersion ($\phi > 1$). While for $\phi = 1$, the standard Poisson model results.

**B Proof of dispersion for DE case**

**Theorem 1.** Let $Y_i, i = 1, \ldots, n$, be (type 1) DE distributed, with $\lambda > 0$. Then, the distribution only allows for overdispersion and equidispersion (when $\lambda \to +\infty$).

**Proof.** To prove it, three situations are examined:

1. $E(Y_i) > \text{Var}(Y_i)$?
   
   \[
   \frac{e^{-\lambda}}{(1 - e^{-\lambda})} > \frac{e^{-\lambda}}{(1 - e^{-\lambda})^2} \iff e^{-\lambda} < 0 \quad \Rightarrow \bot
   \]

2. $E(Y_i) = \text{Var}(Y_i)$?
   
   \[
   e^{-\lambda} = \frac{e^{-\lambda}}{(1 - e^{-\lambda})^2} \iff e^{-\lambda} = 0 \quad \Rightarrow \lambda \to +\infty
   \]

3. $E(Y_i) < \text{Var}(Y_i)$?
   
   \[
   \frac{e^{-\lambda}}{(1 - e^{-\lambda})} < \frac{e^{-\lambda}}{(1 - e^{-\lambda})^2} \iff e^{-\lambda} > 0 \quad \Rightarrow \sqrt
   \]

Thus, the distribution only allows for overdispersion and equidispersion (when $\lambda \to +\infty$)!

**C Proof of mean and variance convergence for DW case**

**Lemma 2 (d’Alembert’s ratio test).** Let $\sum_{n=0}^{+\infty} a_n$ be an infinite serie, and consider

\[
L = \lim_{n \to +\infty} \left| \frac{a_{n+1}}{a_n} \right|.
\]
1. If $L < 1$, then the series converges absolutely;
2. If $L > 1$, then the series diverges;
3. If $L = 1$ or the limit fails to exist, then the test is inconclusive.

**Lemma 3 (Raabe–Duhamel’s test).** Let $a_n > 0 \ (\forall n)$. Define

$$b_n = n \cdot \left( \frac{a_n}{a_{n+1}} - 1 \right).$$

If $L = \lim_{n \to +\infty} b_n$ exists, there are three possibilities:

1. If $L > 1$, then the series $\sum_{n=0}^{+\infty} a_n$ converges;
2. If $L < 1$, then the series $\sum_{n=0}^{+\infty} a_n$ diverges;
3. If $L = 1$, then the test is inconclusive.

**Theorem 4.** Let $Y_i, i = 1, \ldots, n$, be (type 1) DW distributed. Then, it can be shown that

$$E(Y_i) \left( = \mu = \sum_{n=1}^{+\infty} q^n \right) < +\infty,$$

$$\text{Var}(Y_i) \left( = 2 \cdot \sum_{n=1}^{+\infty} n \cdot q^n - \mu - \mu^2 \right) < +\infty.$$  

**Proof.** A trivial proof can be conducted for $\rho \geq 1$, since $\sum_{n=0}^{+\infty} q^n \leq \sum_{n=0}^{+\infty} q^n = (1 - q)^{-1}$, $\sum_{n=0}^{+\infty} n \cdot q^n \leq \sum_{n=0}^{+\infty} n \cdot q^n$, and the series $\sum_{n=0}^{+\infty} n \cdot q^n$ converges. Indeed, using Lemma 2 with $a_n = n \cdot q^n$, it can easily be shown that $L < 1$. For $\rho < 1$, Lemma 3 can be used, where $a_n = n \cdot [q^n - q^{(n+1)}]$ and $a_n = n^2 \cdot [q^n - q^{(n+1)}]$ are proper choices for $E(Y_i)$ and $\text{Var}(Y_i)$, respectively. \[\square\]

Additionally, based on the integral test (Knopp, 1951) and assuming $q = e^{-\lambda}$, the following lower and upper boundaries can be obtained for the mean and variance expression ($\forall t \in 1,2,\ldots$):
A Weibull-count approach for under- and overdispersed longitudinal/clustered data

\[ E(Y_i) \in \left[ \sum_{n=1}^{t} q^{n_\rho} + \int_{t+1}^{+\infty} q^{n_\rho} \cdot dn; \sum_{n=1}^{t} q^{n_\rho} + \int_{t}^{+\infty} q^{n_\rho} \cdot dn \right] \]

\[ \in \left[ \sum_{n=1}^{t} e^{-\lambda^{n_\rho}} + \frac{1}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma[1/\rho; \lambda \cdot (t + 1)^\rho]; \sum_{n=1}^{t} e^{-\lambda^{n_\rho}} + \frac{1}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma(1/\rho; \lambda \cdot t^\rho) \right], \]

\[ \text{Var}(Y_i) \in \left[ 2 \cdot \sum_{n=1}^{t} n \cdot q^{n_\rho} - \sum_{n=1}^{t} q^{n_\rho} - \left( \sum_{n=1}^{t} q^{n_\rho} \right)^2 - \frac{2}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma(1/\rho; \lambda \cdot t^\rho) \cdot \sum_{n=1}^{t} q^{n_\rho} - \frac{1}{\rho^2 \cdot \lambda^{2/\rho}} \cdot [\Gamma(1/\rho; \lambda \cdot t^\rho)]^2 - \frac{1}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma(1/\rho; \lambda \cdot t^\rho) + \frac{2}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma[2/\rho; \lambda \cdot (t + 1)^\rho]; \right. \]

\[ 2 \cdot \sum_{n=1}^{t} n \cdot q^{n_\rho} - \sum_{n=1}^{t} q^{n_\rho} - \left( \sum_{n=1}^{t} q^{n_\rho} \right)^2 - \frac{2}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma[1/\rho; \lambda \cdot (t + 1)^\rho] \cdot \sum_{n=1}^{t} q^{n_\rho} \]

\[ - \frac{1}{\rho^2 \cdot \lambda^{2/\rho}} \cdot \left\{ \Gamma[1/\rho; \lambda \cdot (t + 1)^\rho] \right\}^2 - \frac{1}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma[1/\rho; \lambda \cdot (t + 1)^\rho] + \frac{2}{\rho \cdot \lambda^{1/\rho}} \cdot \Gamma(2/\rho; \lambda \cdot t^\rho) \right]. \]

D SAS code for the hierarchical DE and DW model

/****************************
SOFTWARE: SAS 9.4.
OBJECTIVE: Analyzing Moerzeke data with the DE and DW approach;
DATASET: Moerzeke data, containing information about 457 families;
VARIABLE DESCRIPTION:
− ID: Family ID;
− FamilyMember: Family member indicator, i.e., F = father, M = mother, C = first born child;
− Sexnum: Binary indicator of the gender of first born child, i.e.,
  1 = boy, 0 = girl;
− y: Discretised life expectancy of a household member;
AUTHOR: M. Luyts (L-Biostat);
/****************************

libname m 'C:\Users\u0106491\Desktop\Moerzeke data';
/* Model from exponential case, via NLMIXED */
proc nlmixed data = m.MoerzekeFinal;
   parms beta0=-1 beta0X=-0.4 beta0XX=-0.4 beta1=-0.0134
        beta1X=-0.018 beta1XX=-0.018 sigma=1;
   if FamilyMember='F' then
      eta=beta0XX + beta1XX*SexNum + u;
   else if FamilyMember='M' then
      eta=beta0X + beta1X*SexNum + u;
   else eta=beta0 + beta1*SexNum + u;
   expeta = exp(eta);
   ll = eta*y - log(expeta + 1)*y + log(1 - (expeta/(expeta+1)));
   model y ~ general(ll);
   random u ~ normal(0, exp(sigma)**2) subject=id;
   estimate 'random intercept' exp(sigma);
run;

/* Model from Weibull case, via NLMIXED */
proc nlmixed data = m.MoerzekeFinal;
   parms beta0=-1 beta0X=-0.4 beta0XX=-0.4 beta1=-0.0134
        beta1X=-0.018 beta1XX=-0.018 sigma=1 rho=1;
   if FamilyMember='F' then
      eta=beta0XX + beta1XX*SexNum + u;
   else if FamilyMember='M' then
      eta=beta0X + beta1X*SexNum + u;
   else eta=beta0 + beta1*SexNum + u;
   lambda = log(exp(eta)+1) - log(exp(eta));
   if y=0 then prob = 1 - exp(-1*lambda);
   else prob=exp(-1*lambda*(y**rho)) - exp(-1*lambda*((y+1)**rho));
   ll = log(prob);
   model y ~ general(ll);
   random u ~ normal(0, exp(sigma)**2) subject=id;
   estimate 'random intercept' exp(sigma);
run;
E Characteristic indices for the COM, DP and NB

In spirit of Section 3, similar characteristic indices w.r.t. the Poisson distribution can be obtained for the COM, DP and NB distributions.

1 Negative binomial distribution

![Figure 7: Characteristic indices of NB related to the Poisson distribution.](image)
2 Conway-Maxwell-Poisson distribution

Figure 8: Characteristic indices of COM related to the Poisson distribution.
3 Double Poisson distribution

Figure 9: Characteristic indices of DP related to the Poisson distribution.